



A Galaxy Image Augmentation Method Based on Few-shot Learning and Generative Adversarial Networks

Yiqi Yao¹, Jinqu Zhang² , Ping Du³, and Shuyu Dong²

¹ School of Artificial Intelligence, South China Normal University, Foshan 528225, China

² School of Computer Science, South China Normal University, Guangzhou 510631, China; zjq@scnu.edu.cn

³ Guangdong Construction Vocational Technology Institute, Guangzhou 511500, China

Received 2023 October 9; revised 2024 January 2; accepted 2024 January 15; published 2024 March 4

Abstract

Galaxy morphology classifications based on machine learning are a typical technique to handle enormous amounts of astronomical observation data, but the key challenge is how to provide enough training data for the machine learning models. Therefore this article proposes an image data augmentation method that combines few-shot learning and generative adversarial networks. The Galaxy10 DECaLS data set is selected for the experiments with consistency, variance, and augmentation effects being evaluated. Three popular networks, including AlexNet, VGG, and ResNet, are used as examples to study the effectiveness of different augmentation methods on galaxy morphology classifications. Experiment results show that the proposed method can generate galaxy images and can be used for expanding the classification model's training set. According to comparative studies, the best enhancement effect on model performance is obtained by generating a data set that is 0.5–1 time larger than the original data set. Meanwhile, different augmentation strategies have considerably varied effects on different types of galaxies. FSL-GAN achieved the best classification performance on the ResNet network for In-between Round Smooth Galaxies and Unbarred Loose Spiral Galaxies, with F1 Scores of 89.54% and 63.18%, respectively. Experimental comparison reveals that various data augmentation techniques have varied effects on different categories of galaxy morphology and machine learning models. Finally, the best augmentation strategies for each galaxy category are suggested.

Key words: techniques: image processing – galaxies: structure – galaxies: general











1. Introduction

The morphology of galaxies reflects their internal structure and motion, which can be used to explain the formation and evolution of galaxies, and consequently the history of the universe's formation (Holmberg 1958; Roberts & Haynes 1994; Allen et al. 2006; Benson 2010; Conselice 2014). Astronomers could only categorize galaxies visually in the beginning due to the limitations of observational conditions and techniques, which is known as the visual classification system (Peng et al. 2002; Lotz et al. 2004; Kartaltepe et al. 2015). The most widely used classification scheme for galaxy morphology is the “Hubble sequence” proposed in 1926, which classifies galaxy morphology into three groups: elliptical galaxies, spiral galaxies, and irregular galaxies (Hubble 1926). Nowadays, the era of large field sky surveys has arrived as a result of advancements in observational equipment. The amount of data collected for observing galaxies has increased exponentially since the advent and development of surveys like the Legacy Survey of Space and Time (LSST; Abell et al. 2009), Euclid (Laureijs et al. 2011) and Chinese Survey Space Telescope (CSST; Zhan 2021). It is no longer possible to manage the enormous amounts of data obtained from large field sky

surveys by relying simply on human resources. Even for the Galaxy Zoo project, which has widespread participation, there have only been a few hundred thousand accurate classifications of galaxies done (Willett et al. 2013). Therefore, machine learning-based automatic classification of galaxy morphology is now regarded as the primary way for dealing with enormous amounts of data.

Numerous studies have investigated various machine learning techniques and methodologies for categorizing galaxies based on their morphological properties. Both supervised and unsupervised methods are employed, with supervised methods being the most common. In 2015, for example, a deep neural network model was trained for galaxy morphology classification, and it won an international competition called the Galaxy Challenge (Dieleman et al. 2015). Aniyani & Thorat (2017) presented CNNs for radio galaxies classification, and Vega-Ferrero et al. (2021) investigated the categorization of distant galaxies in the Dark Energy Survey using a supervised deep learning method. Gupta et al. (2022) created a continuous depth version of ResNet to categorize galaxy images from the Galaxy Zoo2 data set. Zhang et al. (2022) originally demonstrated a combination of few-shot learning and CNN for galaxy morphology categorization. It can

Table 1
Basic Information for Each Class in Galaxy10 DECaLS

Sample image	Info.	Sample image	Info.
	Code: 1 Type: Disturbed Galaxies Number: 1081		Code: 6 Type: Barred Spiral Galaxies Number: 2043
	Code: 2 Type: Merging Galaxies Number: 1853		Code: 7 Type: Unbarred Tight Spiral Galaxies Number: 1829
	Code: 3 Type: Round Smooth Galaxies Number: 2645		Code: 8 Type: Unbarred Loose Spiral Galaxies Number: 2628
	Code: 4 Type: In-between Round Smooth Galaxies Number: 2027		Code: 9 Type: Edge-on Galaxies without Bulge Number: 1423
	Code: 5 Type: Cigar Shaped Smooth Galaxies Number: 334		Code: 10 Type: Edge-on Galaxies with Bulge Number: 1873

be concluded that the primary machine learning technique is significantly influenced by the training sample data. Typically, the classification model's accuracy increases with the amount of training data. As a result, increasing the quantity of training samples through data augmentation has emerged as a key strategy for improving model performances.

Data augmentation for galaxy morphology classification is mostly based on image data augmentation technology, which entails fine-tuning existing images or generating new composite images based on existing images to increase training samples. According to the operation of image data augmentation, it can be divided into three methods, including the single image

transformation (SIT) method, multiple images synthesis method, and deep learning-based image generation method. The single image transformation method creates a new image by applying straightforward geometric or color adjustments to a single input image. Image flipping, cropping, scaling, rotation, translation, and other geometric operations are commonly used. Color dithering, histogram equalization, kernel filter sharpening, image blurring, random erasure, etc. are examples of image color modifications. Image geometry adjusting is now the most often utilized method for augmenting galaxy images. For example, Dieleman et al. (2015) used five random perturbations to each training sample during galaxy morphology classification, including randomly

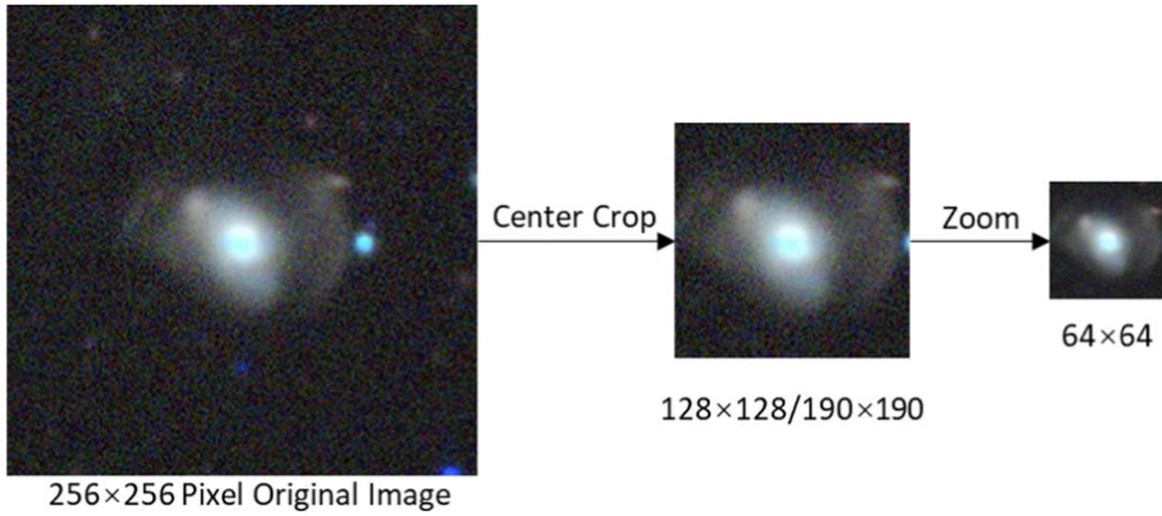


Figure 1. The process of preprocessing.

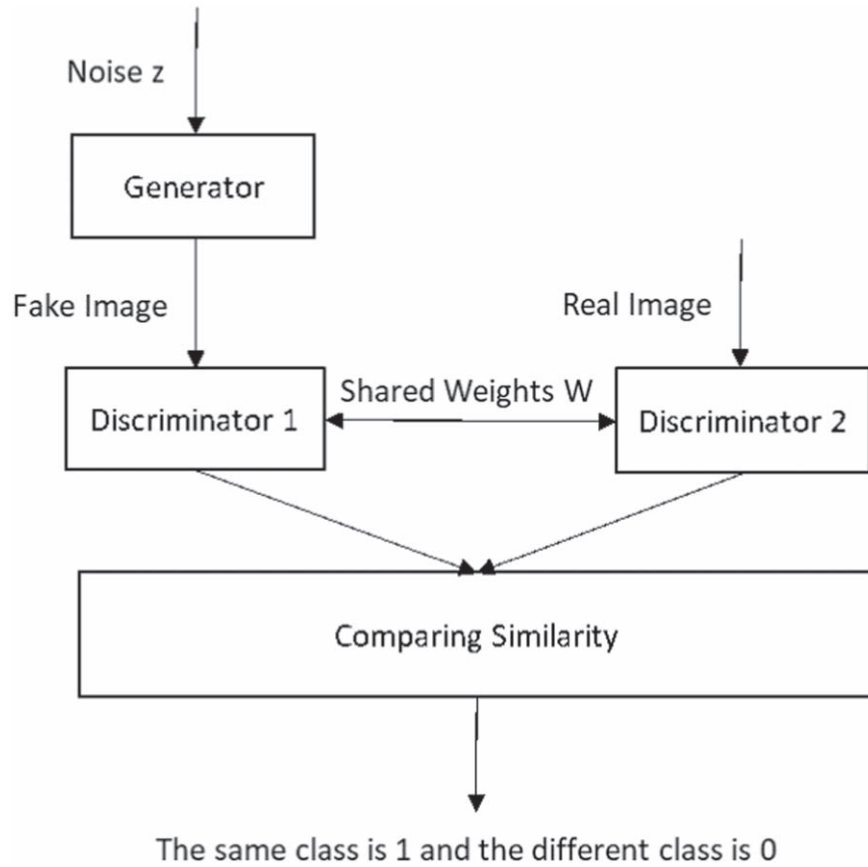


Figure 2. The overall framework of the method.

rotating, shifting, resizing, flipping the image, and modify the brightness of the image. Kim & Brunner (2016) increased the data set size by rotating, flipping, cropping, and introducing Gaussian noise before training the Star-galaxy Classification. Cavanagh et al.

(2021) used cropping, rotating, and flipping to increase the number of data sets by a factor of 40 somewhat strengthening the spatial and translational invariance of the galaxy classification model. These operations significantly increase the amount of training

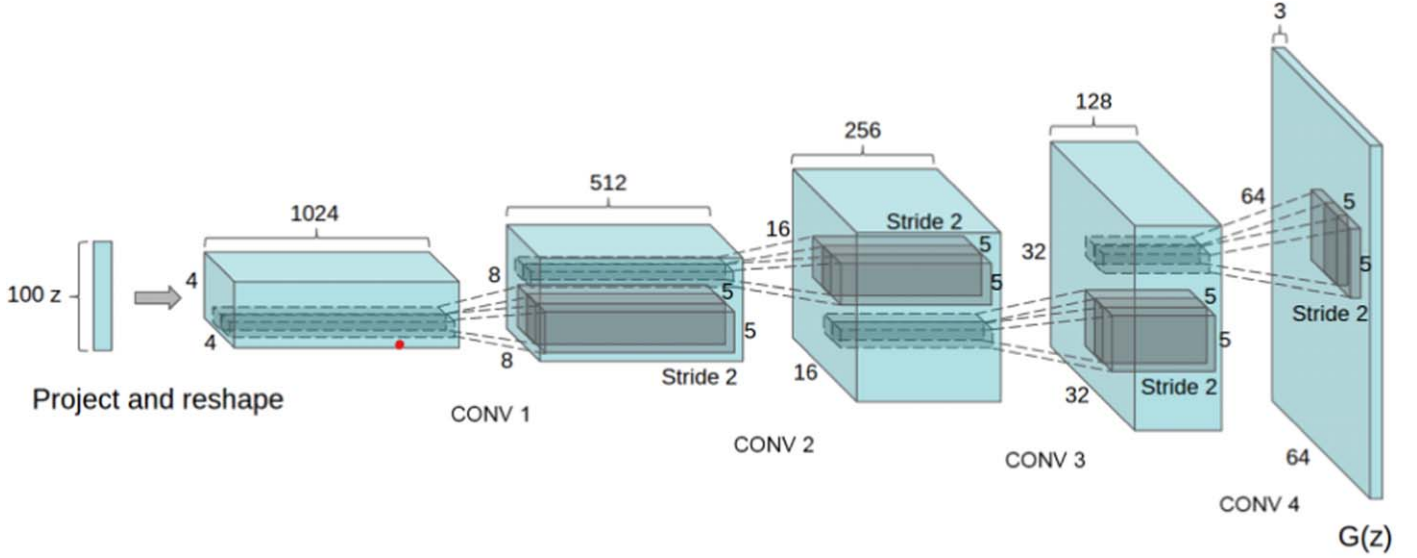


Figure 3. The structure of the generator.

Table 2

The Structure, Output Shape and Number of Parameters of the Generator

Layer	Output Shape	Param
Input	$1 \times 1 \times 100$	0
ConvTranspose2d	$4 \times 4 \times 512$	819,200
BatchNorm2d	$4 \times 4 \times 512$	1,024
ReLU	$4 \times 4 \times 512$	0
ConvTranspose2d	$8 \times 8 \times 256$	2,097,152
BatchNorm2d	$8 \times 8 \times 256$	512
ReLU	$8 \times 8 \times 256$	0
ConvTranspose2d	$16 \times 16 \times 128$	524,288
BatchNorm2d	$16 \times 16 \times 128$	256
ReLU	$16 \times 16 \times 128$	0
ConvTranspose2d	$32 \times 32 \times 64$	131,072
BatchNorm2d	$32 \times 32 \times 64$	128
ReLU	$32 \times 32 \times 64$	0
ConvTranspose2d	$64 \times 64 \times 3$	3,072
Tanh	$64 \times 64 \times 3$	0

samples and improve the model's accuracy. The single image transformation method is simple to use, has a clear meaning, and is executed directly on the image's pixels. Image manipulation, on the other hand, may change the image's content structure, resulting in enhancement failure.

Multiple image synthesis uses two or more images of the same type at the same time to synthesize certain image samples via intermediate interpolation or image mixing. For example, Inoue (2018) developed the SamplePairing approach for image data augmentation, which randomly takes two images from the training set and then adds and averages the two images to create a new training image. Yun et al. (2019) CutMix approach generates new training data by cutting and exchanges a piece of two randomly selected images. The Mosaic image

augmentation method proposed by Bochkovskiy et al. (2020) in YOLO V4 involves randomly selecting four images for cropping and then stitching them together. Mixup and other techniques are additional methods that are a part of multiple image synthesis techniques (Zhang et al. 2017a). This kind of augmentation can readily destroy the brightness distribution of galaxy images, resulting in the formation of illogical images.

The deep learning-based method is a way of autonomously generating new images based on defined feature parameters after learning the features of known training examples. The primary technical tool for this kind of strategy is the use of Generative Adversarial Networks (GAN) to create images (Goodfellow et al. 2014). GAN-based augmentation method has achieved good results on tasks such as image segmentation (Chaitanya et al. 2021), face recognition (Antipov et al. 2017; Huang et al. 2017), and image inpainting (Yeh et al. 2017; Demir & Unal 2018). Given the excellent characteristics of GAN, it has also been used to generate galaxy images. For example, Ravanbakhsh et al. (2017) trained a conditional VAE (Kingma & Welling 2013) and a conditional GAN to create galaxy images. Fussell & Moews (2019) investigated the application of the StackGAN model (Zhang et al. 2017b) for modeling images of high resolution in more detail. According to the principle of generative adversarial networks, it can be known that the more training images used for generator learning, the better its generation result. Because labeled data for galaxy morphology images is both scarce and imbalanced distributed, the main problem that needs to be solved immediately is how to enable GAN models to learn from a small number of samples before applying them to increase the number of galaxy images and improve the classification result.

This study, inspired by Zhang et al. (2022) who introduced few-shot learning to galaxy classification, aims to construct a

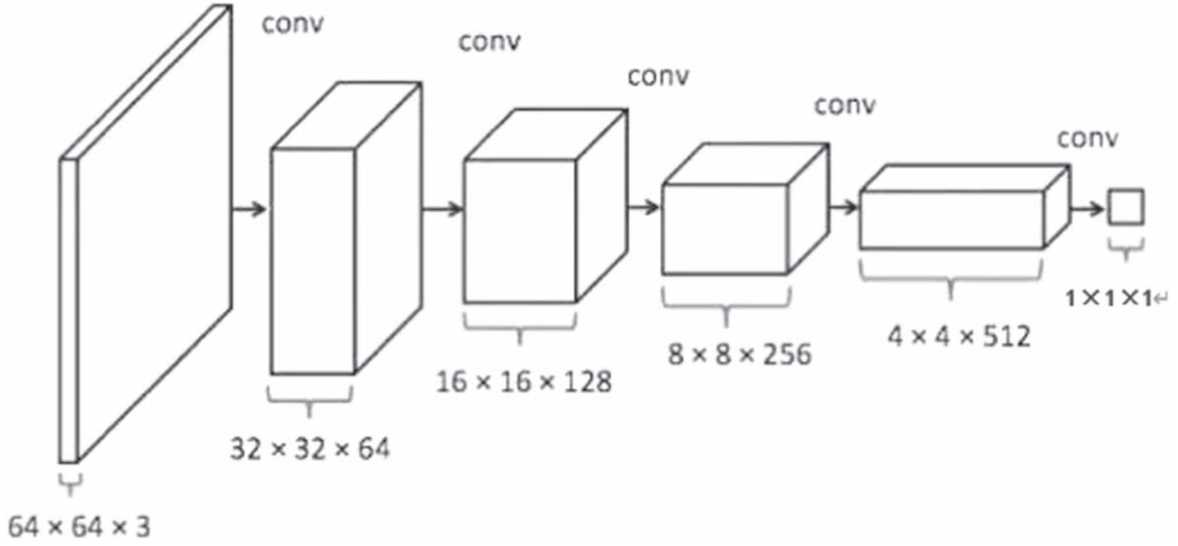


Figure 4. The structure of the generator.

Table 3

The Structure, Output Shape and Number of Parameters of the Discriminator

Layer	Output Shape	Param
Input	$64 \times 64 \times 3$	0
Conv2d	$32 \times 32 \times 64$	3,072
LeakyReLU	$32 \times 32 \times 64$	0
Conv2d	$16 \times 16 \times 128$	131,072
BatchNorm2d	$16 \times 16 \times 128$	256
LeakyReLU	$16 \times 16 \times 128$	0
Conv2d	$8 \times 8 \times 256$	524,288
BatchNorm2d	$8 \times 8 \times 256$	512
LeakyReLU	$8 \times 8 \times 256$	0
Conv2d	$4 \times 4 \times 512$	2,097,152
BatchNorm2d	$4 \times 4 \times 512$	1,024
LeakyReLU	$4 \times 4 \times 512$	0
Conv2d	$1 \times 1 \times 1$	8,192

Few-shot Learning based GAN (FSL-GAN) method that can produce high-quality galaxy images from a limited number of labeled galaxy images. The proposed method can be used to generate a variety of galaxy images that can then be utilized as training data for galaxy morphology classification algorithms, boosting the models' efficiency.

The remainder of the paper is organized as follows: The data set utilized in the study was explained in Section 2. Section 3 introduces the Few-shot Learning-based GAN approach and evaluation. Section 4 contains the results and discussions, while Section 5 has the final conclusions.

2. Dataset

The existing data set employed for machine learning classification of galaxy morphology primarily originates from

Galaxy Zoo. In this article, we leverage the Galaxy 10 DECaLS data set,⁴ a subset produced within the broader Galaxy Zoo project.

2.1. Galaxy10 DECaLS

Images for the Galaxy10 DECaLS were obtained from DESI Legacy Imaging Surveys, and labels were obtained from Galaxy Zoo. It is an improved version of the original Galaxy10 data set that includes Galaxy Zoo Data Release 2 (GZ DR2, Lintott et al. 2011) of DESI Legacy Imaging Surveys (DECaLS, Dey et al. 2019) images rather than Sloan Digital Sky Survey (SDSS, York et al. 2000) images (Walmsley et al. 2022). In this collection, 441,000 galaxies were chosen, and 17,736 of these galaxies were then divided into 10 groups by volunteers. Table 1 displays the Galaxy10 DECaLS code, name, number, and original galaxy morphology image for each class.

2.2. Data Preprocessing

Each galaxy image in the data set was extracted from large field sky survey images. The major galaxy is in the center of the image, surrounded by a dark background, but there is a lot of noise in the background, so the image can be cropped based on the center of the image, which decreases both the noise of the input image and the amount of model calculation. The main part of the edge-on galaxy image in the data set is flatter and longer than that of other classes of galaxy images, and most of the main part of the edge-on galaxy is in the image's diagonal position, so if the clipping range is too large, the main part of the edge-on galaxy may be cut off and important features may

⁴ <https://astronn.readthedocs.io/en/latest/galaxy10.html#>

Table 4
Quality of Generated Images

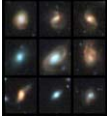
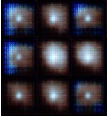
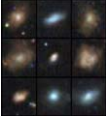
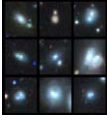
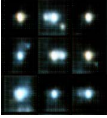
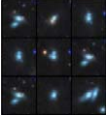
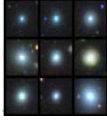
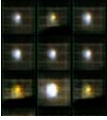

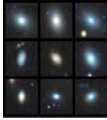
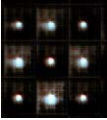
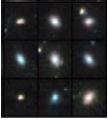

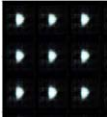

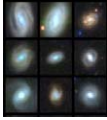
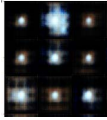
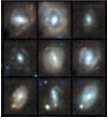
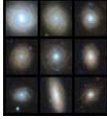
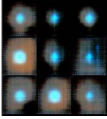
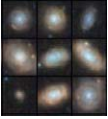
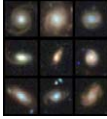
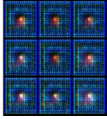
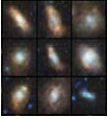

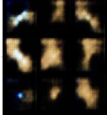
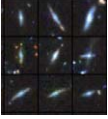

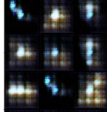

Code	Number of training sets	Original images	DCGAN generated images	FSL-GAN generated images
1	1081			
2	1853			
3	2645			
4	2027			
5	334			
6	2043			
7	1829			
8	2628			
9	1423			
10	1873			

Table 5
The Accuracy of Original and Generated Images

Code	Galaxy Type	Accuracy (%)	
		Original	Generated
1	Disturbed Galaxies	98.1	100
2	Merging Galaxies	99.7	100
3	Round Smooth Galaxies	99.3	100
4	In-between Round Smooth Galaxies	99	100
5	Cigar Shaped Smooth Galaxies	100	100
6	Barred Spiral Galaxies	98.3	99.3
7	Unbarred Tight Spiral Galaxies	98.7	97.6
8	Unbarred Loose Spiral Galaxies	94.1	95.89
9	Edge-on Galaxies without Bulge	99.8	99.6
10	Edge-on Galaxies with Bulge	99	100

be lost. Therefore, the 256×256 pixels edge-on galaxy image is center cropped to 190×190 , while the other categories are center cropped from 256×256 pixels to 128×128 , respectively. To successfully save computational resources, the cropped images are all reduced to the size of 64×64 pixels (Figure 1). We chose 64 as the length and width of each of our layers because there are already many literatures proving that the 64×64 input meets the requirements, so this article directly refers to commonly used sizes (Radford et al. 2015).

3. Methods

In this section, a detailed introduction to the implementation of few-shot learning-based GAN will be provided. Concurrently, an explanation of the evaluation technique will be given.

3.1. Design of Few-shot Learning Based GAN

This paper proposes the idea of few-shot learning using two discriminators with shared weights based on the Deep Convolutional Generative Adversarial Networks (DCGAN) developed by Radford et al. (2015). DCGAN's discriminator requires a large number of known samples to train its discriminative ability. However, galaxy images lack sufficient training samples. Taking Galaxy10 DECaLS data set for example, there are 2645 Round Smooth Galaxies with the most categories, while Cigar Shaped Smooth Galaxies only have 334 samples (Walmsley et al. 2022). As a result, we incorporate few-shot learning methodologies into DCGAN. It is decided whether the two images are of the same class by comparing the similarity of the two images inputted into the discriminators. Figure 2 depicts the overall framework of the method. It consists of three different parts: Generator, Discriminator, and Similarity comparison.

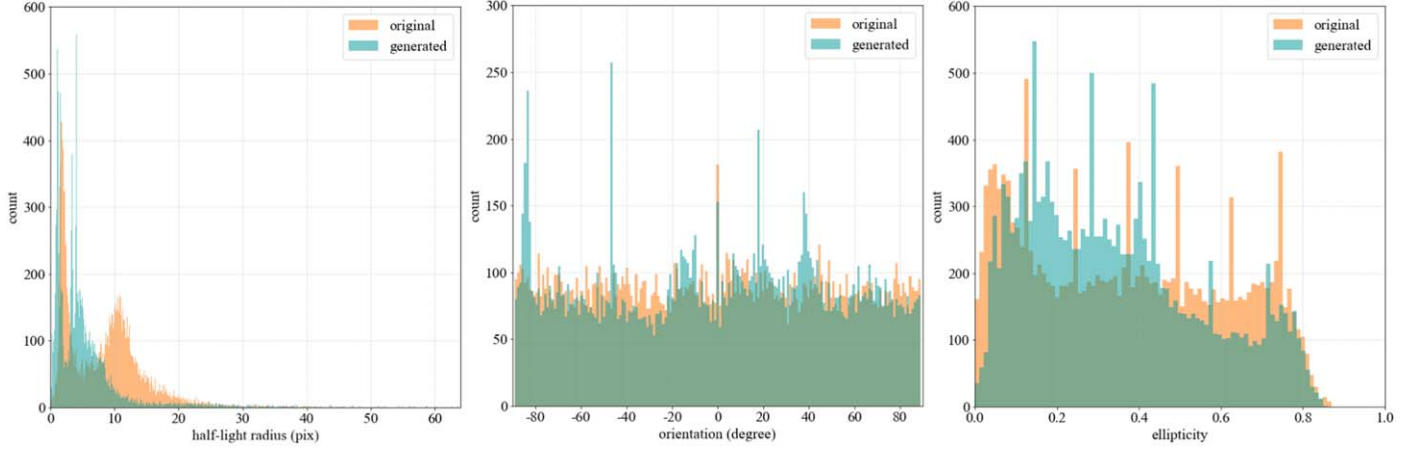


Figure 5. Distribution maps of half-light radius, orientations, ellipticity for original and generated galaxy images.

Table 6
Accuracy, Precision, Recall, and F1-score of Different Models with Highest Values in Bold

Model	Number of Images			Metrics (%)			
	Real Images	Generated Images	Total	Accuracy	Precision	Recall	F1-score
AlexNet	17 736	0	17 736	65.5	65.75	65.5	64.87
	17 736	8864	26 600	66.18	65.73	66.18	65.15
	17 736	17 734	35 470	69.84	69.78	69.84	69.04
	17 736	26 604	44 340	68.56	68.26	68.56	67.49
	17 736	35 464	53 200	68.16	67.3	68.16	67.11
VGG	17 736	0	17 736	68.43	67.91	68.43	66.72
	17 736	8864	26 600	70.69	70.4	70.76	70.55
	17 736	17 734	35 470	72.6	71.66	72.63	71.77
	17 736	26 604	44 340	68.6	68.68	68.61	68.56
	17 736	35 464	53 200	67.03	67.77	67.02	67.28
ResNet	17 736	0	17 736	74.86	74.18	74.86	74.22
	17 736	8864	26 600	76.81	76.37	76.81	76.51
	17 736	17 734	35 470	76.04	76.14	76.07	76.05
	17 736	26 604	44 340	74.87	74.56	74.9	74.7
	17 736	35 464	53 200	72.82	72.34	72.86	72.55

3.1.1. Structure of Generator

The generator accepts a 100-dimensional random noise z that follows the normal distribution, and through the process of deconvolution, transforms it into a $64 \times 64 \times 3$ RGB image. Figure 3 and Table 2 both show the number of parameters for each layer of the generator. The structure uses five deconvolutional layers for up-sampling. Each layer, with the exception of the output layer, is batch-normalized to prevent deviations caused on by an increase in the number of layers. The last layer employs the Tanh activation function, whereas the other layers employ the ReLU activation function (Nair & Hinton 2010).

3.1.2. Structure of Discriminator

In DCGAN, the discriminator should decide if an image is real or a fake created by the generator. Here, the sigmoid activation function of the DCGAN discriminator's last layer is eliminated, and five convolutional layers are employed instead. Figure 4 and Table 3 show the network structure of the discriminator. The discriminator's input consists of two images, X_1 and X_2 , as well as a label Y . If the two images are in the same category, label Y is 1, otherwise it is 0. To avoid bias induced by increasing the number of layers, batch normalization is performed on each layer except the input layer,

Table 7
F1-score of Three Augmentation Techniques for Different Galaxy Categories and Models with Highest Values in Bold

Galaxy Type	Model	Augmentation Method		
		FSL-GAN	SIT	Hybrid
Disturbed Galaxies	AlexNet	33.33	46.26	46.12
	VGG	28.48	53.88	43.04
	ResNet	32.54	56.35	49.04
Merging Galaxies	AlexNet	67.27	75.59	76.13
	VGG	71.56	81.89	79.3
	ResNet	80.74	76.98	79.77
Round Smooth Galaxies	AlexNet	86.34	85.46	87.5
	VGG	88.73	89.02	89.72
	ResNet	89.5	87.86	88.32
In-between Round Smooth Galaxies	AlexNet	82.57	78.18	82.38
	VGG	84.7	89.2	88.83
	ResNet	89.54	87.27	87.15
Cigar Shaped Smooth Galaxies	AlexNet	57.5	82.53	79.64
	VGG	64.1	92.31	86.72
	ResNet	62.86	91.6	86.73
Barred Spiral Galaxies	AlexNet	65.15	60.69	67.38
	VGG	66.75	79.02	75.54
	ResNet	78.09	81.25	77.49
Unbarred Tight Spiral Galaxies	AlexNet	51.9	63.42	60.06
	VGG	63.08	74.93	71.51
	ResNet	65.17	68.72	65.9
Unbarred Loose Spiral Galaxies	AlexNet	52.44	42.97	51.02
	VGG	56.94	57.22	57.44
	ResNet	63.18	59.03	56.99
Edge-on Galaxies without Bulge	AlexNet	88.22	87.91	85.71
	VGG	84.29	89.69	87.18
	ResNet	86.77	88.67	89.01
Edge-on Galaxies with Bulge	AlexNet	84.12	80.82	82.27
	VGG	85.03	87.84	88.89
	ResNet	87.94	85.97	87.36

forcing the data to follow a fixed data distribution (Ioffe & Szegedy 2015). Meanwhile, the activation function in each layer is Leaky ReLU (Equation (1)), which has a rapid computation speed and can effectively deal with the problem of gradient disappearance (Maas et al. 2013).

$$\text{LeakReLU}(x) = \begin{cases} x, & x > 0 \\ \text{leak} * x, & x \leq 0 \end{cases} \quad (1)$$

3.1.3. Similarity Comparison and Loss Function Design

When training the model, two real images, X1 and X2, are chosen at random from the training set. First, the discriminator is trained. Two fake images, F1 and F2, are generated by feeding noise data into the generator, and then four groups of data

(X1, X2, 1), (F1, F2, 1), (X1, F1, 0), (X2, F2, 0) are inputted into the discriminator to make the discriminator output as similar to the actual situation as possible. Second, the generator is trained. Input noise data to the generator again to make two fake images F3 and F4, and then input two sets of data (X1, F3, 1) and (X2, F4, 1) to the discriminator so that the fake images made by the generator are not recognized by the discriminator as much as possible. Through the interaction between the discriminator and the generator, the discriminator's accuracy is increased while the images produced by the generator become more and more consistent with the distribution law of the real samples, resulting in fake images that are infinitely close to real images. The loss function uses the Comparative Loss function (Hadsell et al. 2006), whose formula is:

$$L = \frac{1}{2N} \sum_1^N YE^2 + (1 - Y)\max(\text{margin} - E, 0)^2 \quad (2)$$

where N represents the number of sample pairs and E represents the Euclidean distance between the two feature vectors output by the discriminator, the Euclidean distance is calculated by Equation (3).

$$E = ||D(X_1) - D(X_2)|| \quad (3)$$

While there is great similarity and a small Euclidean distance between images of the same class, there is a small similarity and a huge Euclidean gap between images of different classes. The threshold margin, which can be changed depending on the actual scenario of the image data, reflects the maximum dissimilarity between images in the same class while still allowing for diversity.

3.2. Evaluation

To verify the usability of images generated by FSL-GAN, the images are evaluated from three perspectives: consistency, variance, and augmentation effects. The feature of a generated image that stays in the same category is called consistency. During the testing phase, we evaluate the generated image's potential to fool the discriminator in order to determine whether it falls into the original category. If the discriminator can be fooled, it indicates that the generated image is consistent with the original category. First, the same number of original images of different classes and generated images of the same class are randomly selected. The pre-trained classifier with original images is then used to categorize two sets of images to see if the classifier can accurately classify the generated images. If the generated galaxies can be classified as accurately as the original ones, it will show that the generated images can successfully overcome the discriminator and fool the classifier. The generated galaxy images will then be eligible for augmentation.

Variance is the existence of distinctions between created and original images. If there is no obvious difference between a generated image and the original image, then it is simply a copy of the original image and cannot be considered as incremental training data. In order to assess the variance of created images, we thus investigate three morphological features of the galaxy

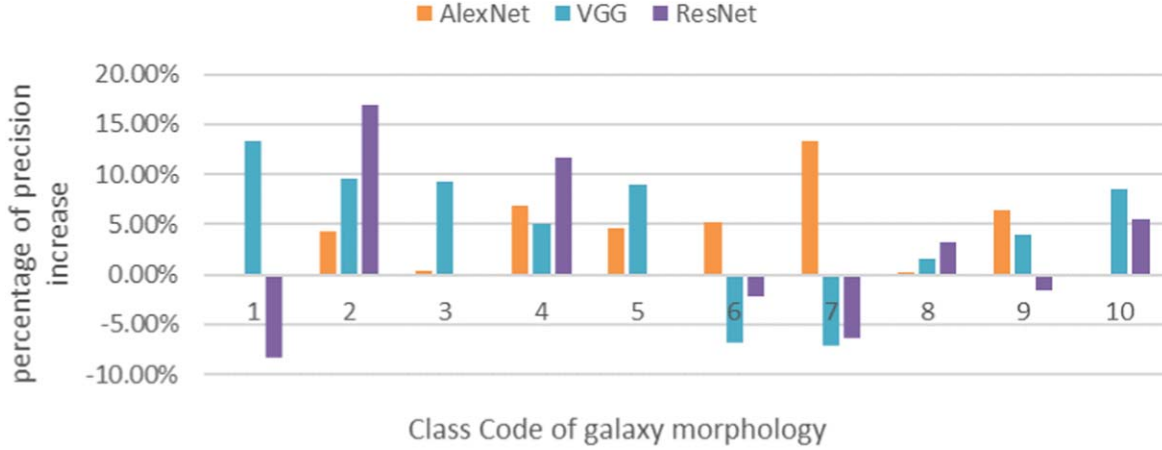


Figure 6. The percentage of precision increase for galaxy morphology classification by using FSL-GAN augmentation on three Network models (The code meaning could be found in Table 1).

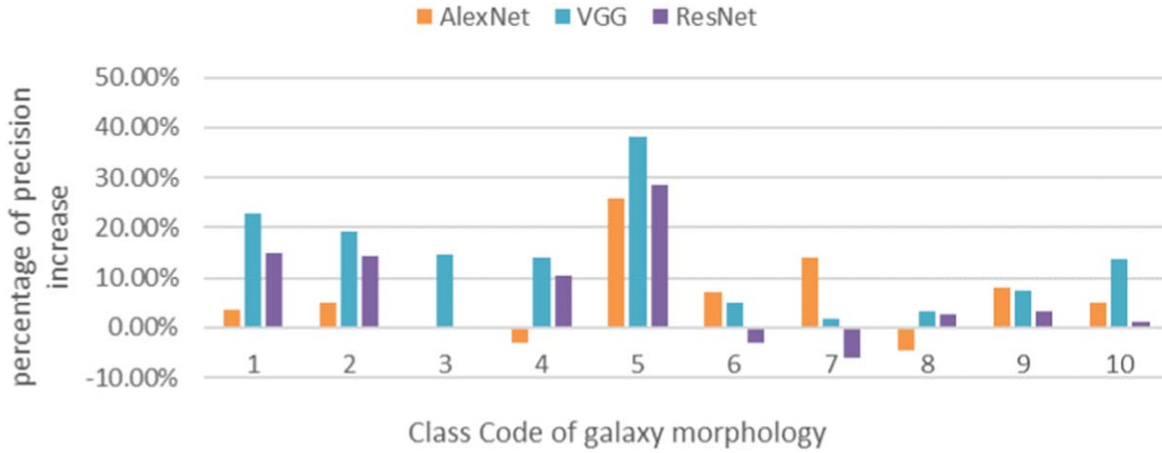


Figure 7. The percentage of precision increase for galaxy morphology classification by using SIT augmentation on three Network models (The code meaning could be found in Table 1).

images: half-light radius, orientations, and ellipticity. The half-light radius is the radius that encloses half of the total flux of the Sérsic profile (Ciotti 1991). The orientation is the angle between the x -axis and the semimajor axis, which increases in a counterclockwise direction. Galaxy ellipticity is the degree to which a galaxy's shape deviates from that of a perfect circle or sphere. It is defined as:

$$\epsilon = 1 - \frac{b}{a} \quad (4)$$

where a and b are the semimajor and semiminor axis of the ellipse respectively. In our study, we use the Photutils package⁵

⁵ <https://photutils.readthedocs.io/en/stable/#>

and Statmorph⁶ package (Rodríguez-Gomez et al. 2019) to calculate these morphological properties.

The augmentation effect refers to the enhancement of classification accuracy achieved by increasing the training set with generated galaxy images by FSL-GAN. Equations (5) through (8) summarize the procedure for calculating the evaluation metrics for the multiple classification task, which include accuracy, precision, recall, and F1-score.

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (5)$$

⁶ <https://statmorph.readthedocs.io/en/latest/index.html>

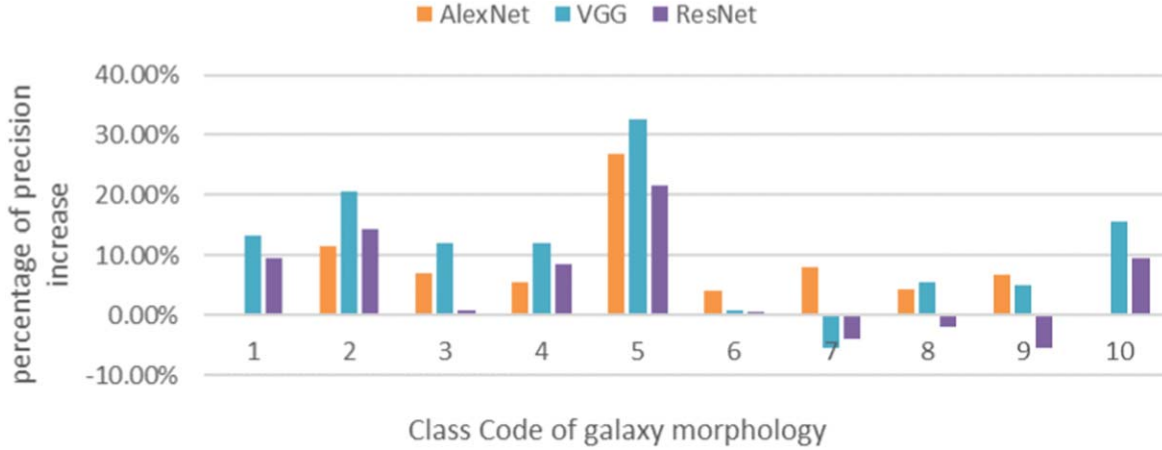


Figure 8. The percentage of precision increase for galaxy morphology classification by using hybrid augmentation on three Network models (The code meaning could be found in Table 1).

$$\text{precision} = \frac{TP}{TP + FP} \quad (6)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (7)$$

$$F1\text{-score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \quad (8)$$

In the equations above, TP stands for true positives, FP for false positives, TN for true negatives, and FN for false negatives.

4. Results and Discussion

We first performed a visualization analysis on images produced by DCGAN and FSL-GAN, followed by a consistency analysis to confirm the usability for augmentation for the images generated by FSL-GAN. Then, we verify the variance of the generated images using metrics such as half-light radius, orientations, and ellipticity. In the end, data augmentation is applied to the galaxy images in order to enhance the model's classification performance.

4.1. Visualization and Consistency Analysis of Generated Images

Both DCGAN and FSL-GAN are trained on the Galaxy10 DECaLS data set, and Table 4 displays the galaxy images generated after 100 epochs. It can be seen from an analysis of the created images of different morphologies that the images generated using the FSL-GAN method are more logical in terms of visual effects.

To verify whether the images generated by FSL-GAN are consistent with the original category, a pre-trained ResNet50 using PyTorch (Paszke et al. 2019) is first constructed and trained using original images. Then, 1000 original images and

1000 generated images are randomly selected for categorization, and the accuracy is given in Table 5. We can see that the accuracy of the generated images is higher than that of the original images in all categories of images except class 7, where it is 1.1% lower, and class 9, where it is 0.2% lower. The results show that the created images have strong consistency with the original category, can effectively trick the discriminator, and can be used to augment the model's training data.

4.2. Variance Analysis of Generated Images

Three metrics are utilized to compare the difference between the generated images and the original images: half-light radius, orientations, and ellipticity. It can be observed from their distribution maps in Figure 5 that the half-light radius, orientations, and ellipticity have the same range but different distributions. Taking half-light radius as an example, there are far more generated images than original ones with a half-light radius greater than 10 lt-yr. The distribution indicates that there is a variation between the generated image and the original image. If the generated images are used for augmentation, it is the same as adding training samples that are not from the original images, so the model's training effect can be enhanced.

4.3. The Impact of Different Numbers of Generated Images on the Classification of Galaxy Morphology

The Galaxy10 DECaLS data set is augmented by the generated images for increasing training samples in order to investigate whether the galaxy images generated by FSL-GAN can be used to improve the performance of the galaxy morphology classification models. We generated 8864, 17,734, 26,604, and 35,464 images, respectively, which are equivalent to 0.5, 1, 1.5, and 2 times the size of the original data set, to determine what quantity of additional images is feasible for the

Table 8

Data Augmentation Strategies for Different Models and Galaxy Morphologies

Code	Name	Model	Augmentation Strategies
1	Disturbed Galaxies	AlexNet	SIT
		VGG	SIT
		ResNet	SIT
2	Merging Galaxies	AlexNet	Hybrid
		VGG	SIT
		ResNet	FSL-GAN
3	Round Smooth Galaxies	AlexNet	Hybrid
		VGG	SIT
		ResNet	FSL-GAN
4	In-between Round Smooth Galaxies	AlexNet	FSL-GAN
		VGG	SIT
		ResNet	FSL-GAN
5	Cigar Shaped Smooth Galaxies	AlexNet	SIT
		VGG	SIT
		ResNet	Hybrid
6	Barred Spiral Galaxies	AlexNet	Hybrid
		VGG	FSL-GAN
		ResNet	Hybrid
7	Unbarred Tight Spiral Galaxies	AlexNet	SIT
		VGG	Hybrid
		ResNet	Hybrid
8	Unbarred Loose Spiral Galaxies	AlexNet	Hybrid
		VGG	Hybrid
		ResNet	SIT
9	Edge-on Galaxies without Bulge	AlexNet	FSL-GAN
		VGG	Hybrid
		ResNet	FSL-GAN
10	Edge-on Galaxies with Bulge	AlexNet	FSL-GAN
		VGG	Hybrid
		ResNet	Hybrid

original data set. Multiple networks, including AlexNet (Krizhevsky 2014), VGG (Simonyan & Zisserman 2015), and ResNet (He et al. 2016), are trained using the extended data set, and the effects of generated images on model training are compared and assessed in Table 6. To ensure comparability, the training and testing sets are divided using the ratio of 9:1, and the samples in the testing set are all real samples from the Galaxy 10 DECaLS data set, with the same number of images in each class. Since it is a multi-classification task, the classification results are assessed using overall accuracy, weighted average precision, weighted average recall, and weighted average F1-score. It can

be noticed from Table 6 that the accuracy, weighted average precision, weighted recall, and weighted F1-score of all models are all improved when the expansion factor is 0.5, 1, 1.5, and 2. However, different models experience different improvements. ResNet has the highest overall accuracy, followed by VGG, while AlexNet has the lowest accuracy. The classification performance of the model is at its optimum when the expansion factor of AlexNet and VGG is one time, while ResNet is 0.5 times. It can be concluded that when the augmentation multiplier reaches a certain level, all the features of the galaxy image have been learned, and further increasing the multiplier will only add some redundant features that will not help or even reduce the classification effect.

4.4. Analysis of the Classification Effect of Data Augmentation on Different Galaxy Morphology Categories

The previous section focused on assessing the overall effect of varied numbers of generated images on galaxy morphology classification, but it is unclear if generated images are effective for all morphological classes. Therefore, we will focus on the classification ability of different augmentation methods on various galaxy categories in this part. We compare the classification performance of three augmentation methods on three network structures. The model parameters are held constant, and the number of augmentation expansions is the optimal multiple for each model as analyzed in the previous section.

Three augmentation techniques—the single image transformation method, the FSL-GAN method, and a hybrid method combining the two—are compared for their performance in classifying galaxy morphologies using the best number of generated images. For the hybrid method, it means generating half of the sample images using the FSL-GAN method and the other half using the single image transformation method.

The F1-score is utilized as an evaluation metric since it considers both precision and recall. Table 7 demonstrates the F1-score results of three augmentation techniques for different galaxy categories and models. It can be noticed that different augmentation methods will impact different galaxy types in different ways. There is no data augmentation method that can produce the best outcomes for all galaxy classifications. For In-between Round Smooth Galaxies and Unbarred Loose Spiral Galaxies on the ResNet model, the FSL-GAN technique has the highest classification F1-score, but the hybrid method achieved the best results on the VGG model for Round Smooth Galaxies and Edge on Galaxies with Bulge. In some galaxy categories, the SIT method performs well for both ResNet and VGG models.

Figures from 6 to 8 further demonstrate the percent increase of precision for galaxy morphology classification by using three different data augmentation methods on each galaxy

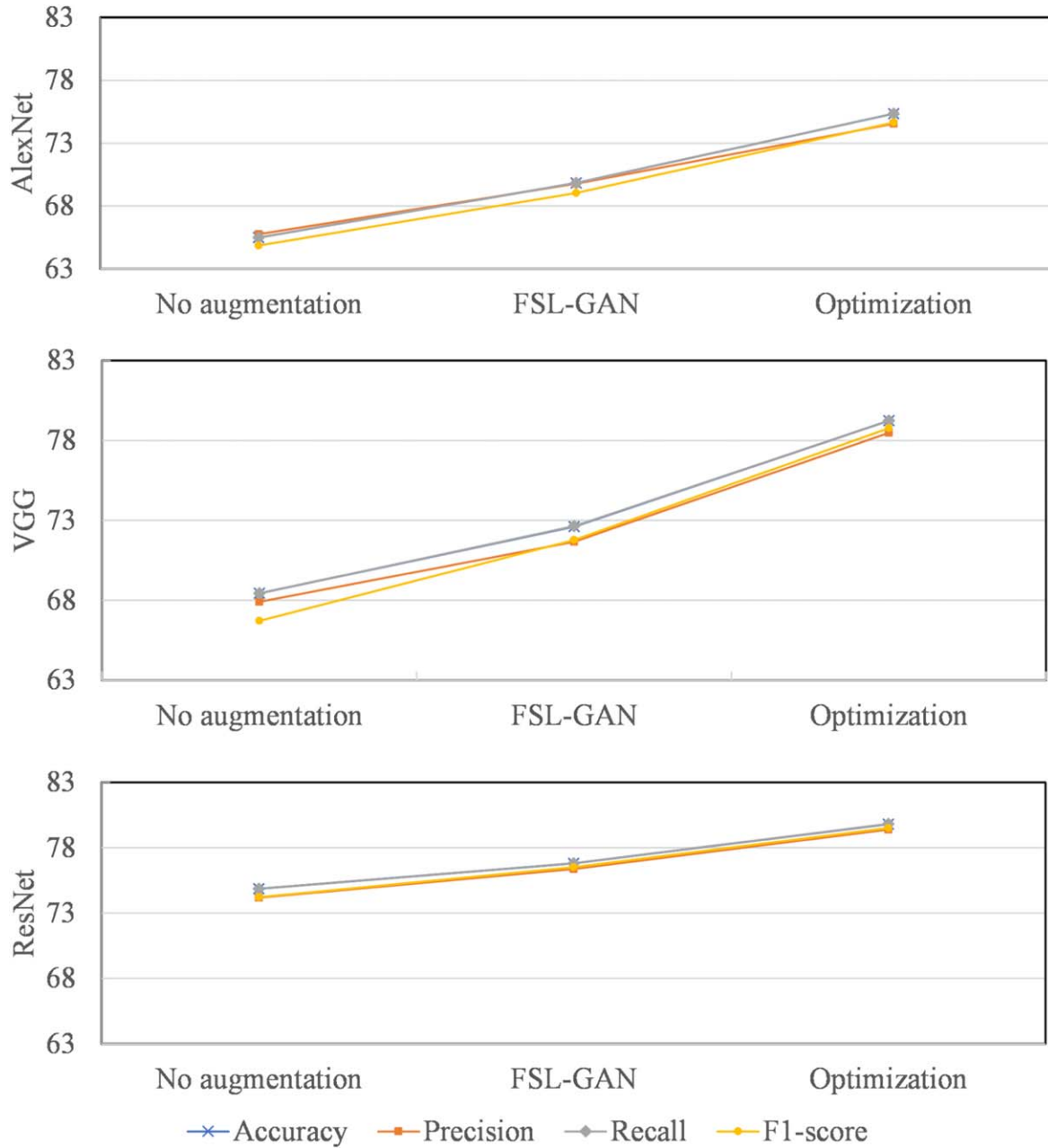


Figure 9. Performance of data augmentation optimization strategies on different models.

category. The three data augmentation approaches considerably enhanced the precision of galaxy morphology classification. However, it can also be seen that data augmentation does not work for some specific categories.

In Figure 6, using AlexNet as an example, FSL-GAN enhanced the classification precision of the seventh category by 13.37% while decreasing the classification precision of the first category by 0.17%. Similarly, instead of boosting the precision of the seventh category when employing FSL-GAN on the

VGG, the precision appears to fall significantly. The classification precision of the second and fourth classes, on the other hand, has significantly improved on each model. Only the second, fourth, and tenth categories on the ResNet showed noticeable improvement, leaving the remaining categories with little change.

From the comparison between Figures 7 and 8, the negative effects of the SIT method on AlexNet found in class 4 and 8 are improved after applying hybrid data augmentation. Regardless of

the augmentation method utilized, class 2 performs better in terms of classification precision. Figure 8 demonstrates the precision improvement of each category under the simultaneous application of hybrid data augmentation. It is observed that the classification precision of class 5 Cigar Shaped Smooth Galaxies has been enhanced the most and achieves the greatest improvement for all three models. It is probably because the number of real images of cigar galaxies is so limited, only 334, so the influence of feature learning is more obvious after data augmentation.

4.5. Data Augmentation Strategies for Different Galaxy Morphologies

The investigation in the previous section led to the conclusion that different models and classifications of galaxy morphology respond differently to various types of data augmentation methods. Therefore, it will be extremely beneficial to each model's overall classification performance by reorganizing the data augmentation technique for each galaxy morphology. In Table 8, we have listed the most appropriate data augmentation strategies suitable for each galaxy morphology for the three distinct models.

Figure 9 illustrates the performance changes and comparisons of three models after applying FSL-GAN data augmentation and optimization strategy augmentation to each galaxy category. The analysis yields that under the same experimental parameters and number of training sets, the accuracy, precision, recall, and F1-score of the three models, including AlexNet, VGG, and ResNet, have remarkably improved. This implies that the FSL-GAN data augmentation method has obvious effect on galaxy morphological classification, but combining it with the single image transformation method to find an optimal augmentation strategy fit for each category can better improve the classification effect.

5. Conclusion

This research proposes a few-shot learning approach with generative adversarial networks as a data augmentation method for galaxy morphology classification. The research first assessed the effectiveness of the images generated by the FSL-GAN approach, and then further developed data augmentation optimization strategies for different categories of galaxy morphology by examining the effects of various data augmentation techniques on them. Three basic conclusions can be drawn from the findings presented in this article:

1. The FSL-GAN data augmentation method is an effective way for improving the performance of galaxy morphology classification, however, it is not applicable to all galaxy morphologies. FSL-GAN achieves the greatest improvement on the ResNet model for categories 2 Merging Galaxies and 4 In-between Round Smooth Galaxies. Since the generator mainly learns upon existing

image features, there are still shortcomings in the variance of the generated images, resulting in a limited effect on certain categories of galaxy morphologies.

2. Different models and different galaxy morphologies have different sensitivities to different data augmentation methods. According to the analysis of the results from three network models, including AlexNet, VGG, and ResNet, no data augmentation method can produce the best outcomes for all galaxy classifications.
3. A key factor in enhancing the performance of galaxy morphology classification is selecting the proper data augmentation methods and related classification models based on different galaxy morphology categories.

With the development of the machine learning theory, more and more models for classifying galaxy morphology have been created. In this article, we utilized three popular models as examples to show how different models learn differently from various galaxy morphologies. More models can be put to the test in actual situations while the capacity to generate various galaxy images will also be continually improved.

Acknowledgments

This work was supported by China Manned Space Program through its Space Application System, the National Natural Science Foundation of China (NSFC, grant Nos. 11 973 022 and U1811464), and the Natural Science Foundation of Guangdong Province (No. 2020A1515010710).

ORCID iDs

Jinqu Zhang  <https://orcid.org/0000-0001-6643-4053>

References

- Abell, P. A., Allison, J., Anderson, S. F., et al. 2009, LSST Science Book FERMILAB-TM-2495-A, LSST Science Collaborations
- Allen, P. D., Driver, S. P., Graham, A. W., et al. 2006, *MNRAS*, **371**, 2
- Aniyan, A., & Thorat, K. 2017, *ApJS*, **230**, 20
- Antipov, G., Baccouche, M., & Dugelay, J.-L. 2017, Face aging with conditional generative adversarial networks, in IEEE International Conference on Image Processing (ICIP) (Piscataway, NJ: IEEE), 2089
- Benson, A. J. 2010, *PhR*, **495**, 33
- Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. 2020, arXiv:2004.10934v1
- Cavanagh, M. K., Bekki, K., & Groves, B. A. 2021, *MNRAS*, **506**, 659
- Chaitanya, K., Karani, N., Baumgartner, C. F., et al. 2021, *Med. Image Anal.*, **68**, 101934
- Ciotti, L. 1991, *A&A*, **249**, 99
- Conselice, C. J. 2014, *ARA&A*, **52**, 291
- Demir, U., & Unal, G. 2018, arXiv:1803.07422
- Dey, A., Schlegel, D. J., Lang, D., et al. 2019, *AJ*, **157**, 168
- Dieleman, S., Willett, K. W., & Dambre, J. 2015, *MNRAS*, **450**, 1441
- Fussell, L., & Moews, B. 2019, *MNRAS*, **485**, 3203
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. 2014, Adv. Neural Inf. Proc. Syst. in Proc. 27th Int. Conf. Neural Inf. Proc. Syst., 2 (Cambridge, MA: MIT Press), 2672
- Gupta, R., Srijith, P., & Desai, S. 2022, *A&C*, **38**, 100543
- Hadsell, R., Chopra, S., & LeCun, Y. 2006, Dimensionality Reduction by Learning an Invariant Mapping, in IEEE Computer Society Conf. on

- Computer Vision and Pattern Recognition (CVPR'06), 2 (New York: IEEE), 1735
- He, K., Zhang, X., Ren, S., & Sun, J. 2016, Deep Residual Learning for Image Recognition, in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (Las Vegas, NV: IEEE), 770
- Holmberg, E. 1958, *Lund Medd. Astron. Obs. Ser. II*, **136**, 1
- Huang, R., Zhang, S., Li, T., & He, R. 2017, Beyond face rotation: global and local perception GAN for photorealistic and identity preserving frontal view synthesis, in Proc. 16th IEEE Int. Conf. Computer Vision (ICCV) (Venice: IEEE), 2458
- Hubble, E. 1926, *ApJ*, **64**, 321
- Inoue, H. 2018, arXiv:1801.02929
- Ioffe, S., & Szegedy, C. 2015, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, in Proc. 32nd Int. Conf. Machine Learn., 37 (Lille France: PMLR), 448
- Kartalpe, J. S., Mozena, M., Kocovski, D., et al. 2015, *ApJS*, **221**, 11
- Kim, E. J., & Brunner, R. J. 2016, *MNRAS*, **464**, 4463
- Kingma, D. P., & Welling, M. 2013, arXiv:1312.6114
- Krizhevsky, A. 2014, arXiv:1404.5997
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
- Lintott, C., Schawinski, K., Bamford, S., et al. 2011, *MNRAS*, **410**, 166
- Lotz, J. M., Primack, J., & Madau, P. 2004, *AJ*, **128**, 163
- Maas, A. L., Hannun, A. Y., Ng, A. Y., et al. 2013, in Proc. 30th Int. Conf. Mach. Learn., 28 (Atlanta, GA: JMLR), 3
- Nair, V., & Hinton, G. E. 2010, Rectified Linear Units Improve Restricted Boltzmann Machines, in Proc. 27th Int. Conf. Mach. Learn. (Haifa: IEEE), 807
- Paszke, A., Gross, S., Massa, F., et al. 2019, in Adv. Neural Inf. Proc. Syst., 32 ed. H. Wallach, H. Larochelle, & A. Beygelzimer (San Jose: Curran Associates, Inc.), 8024
- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, *AJ*, **124**, 266
- Radford, A., Metz, L., & Chintala, S. 2015, arXiv:1511.06434
- Ravanbakhsh, S., Lanusse, F., Mandelbaum, R., Schneider, J., & Poczós, B. 2017, in Proc. of the AAAI Conf. on Artificial Intelligence, 31 (San Francisco, CA: AAAI Press), 1
- Roberts, M. S., & Haynes, M. P. 1994, *ARA&A*, **32**, 115
- Rodriguez-Gomez, V., Snyder, G. F., Lotz, J. M., et al. 2019, *MNRAS*, **483**, 4140
- Simonyan, K., & Zisserman, A. 2015, arXiv:1409.1556
- Vega-Ferrero, J., Domínguez Sánchez, H., Bernardi, M., et al. 2021, *MNRAS*, **506**, 1927
- Walmsley, M., Lintott, C., Geron, T., et al. 2022, *MNRAS*, **509**, 3966
- Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, *MNRAS*, **435**, 2835
- Yeh, R. A., Chen, C., Yian Lim, T., et al. 2017, in IEEE Conf. Computer Vision and Pattern Recognition (CVPR) (Honolulu, HI: IEEE), 5485
- York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, *AJ*, **120**, 1579
- Yun, S., Han, D., Oh, S. J., et al. 2019, in IEEE/CVF Int. Conf. Computer Vision (ICCV) (Seoul: IEEE), 6022
- Zhan, H. 2021, *ChSBu*, **66**, 1290
- Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. 2017a, arXiv e-prints
- Zhang, H., Xu, T., Li, H., et al. 2017, in Proc. IEEE International Conf. Computer Vision (Venice: IEEE), 5907
- Zhang, Z., Zou, Z., Li, N., & Chen, Y. 2022, *RAA*, **22**, 055002